

Should Standardization Tests be Used to Assess Student Ability?

-- An Evaluation on Effects of Standardization Tests

Quanya Yang^{1, *}

¹Dulwich College Suzhou, Suzhou, Jiangsu, 215000, China

*Corresponding author Email: amy.yang23@stu.dulwich.cn

Abstract. Standardization testing, a form of testing that have relatively standard procedure, are being implemented worldwide nowadays to measure student's ability and making academic decisions. Despite its popularity, it remains controversial whether standardization testing is the most appropriate method to assess student. Lots of criticism had been made since the last century, while other researchers think standardization serves a value and there is no replacement for it yet. This paper aims to summarize and evaluate the effect of this implementation on students themselves and the external environment by incorporating opinions proposed by current research so far. Weighting up the of potential merits and drawbacks of using this method of assessment, standardization is still irreplicable while having many negative influences such as supressing creativity and leading to study for the test. Fortunately, many examination boards are taking these influences into account and starting to offer students the freedom to study for their own interests. However, research around this topic is still limited and unspecific. More detailed studies should be conducted to investigate a particular subject and question type; the relationship between students' age group and types standardization testing.

Keywords: Standardization tests, Evaluation, Assessment, Teaching to the test.

1. Introduction

As globalization expands, so as the usage of standardization test. Specifically in the U.S., in 1998, 39 states had adopted some form of standard-based tests or accountability systems (24 of them attaching high stakes in the form of student recognition, promotion, or graduation) , the number had risen to 49 in 2001 [1]. States must bring all students up to the proficient level on state tests by the 2013-2014 school year. But standardization test in 21st century is much more than an indicator of student ability. Tests are used at almost every school worldwide, standardization tests thus affect many lives of children [2]. As Jencks and Phillips [3] put it "it is no longer used as only an indicator of student's ability but also used to decide important academic decision such as college entrance or proceed to the next grade." Therefore, it exists a requirement to examine this educational tool again after its evolvement for a long time. While lots of research are done with both positive and negative comments on it, there is a research gap to summarize all of them and points out some limitations as most of current studies are only focusing on holistic perspective of standardization tests. This paper aims to summarize and analyse on both pros and cons by firstly clarifying definitions and types of standardization tests, then provide suggestions according to current situation of standardization test.

2. Different types of standardization

Standardization tests, according to Lee J Cronbach [4], is which the procedure, apparatus, scoring is fixed so the same test could be given at different times and places. In general, standardization tests are used worldwide for purpose such as school entrance, exit, admission and certification...etc. According to Burkett [5], the majority of tests could be categorized into two types: norm-referenced tests and criterion-referenced test. In norm-referenced tests (NTRs), performances of test takers are compared to other selected takers basing on characteristics such as gender and age. Individual test takers are usually ranked in comparison to others as a percentile group. NRTs are usually used to

assess readiness for school, academic level, examples include SAT and ACT. In criterion-referenced test (CRTs), introduced by American educational psychologists Robert Glaser [6], takers are assessed with a fixed and standard criterion. This type of assessment decides whether the taker have reached a certain goal or master of a skill. Scores are presented in how many questions are answered correctly to the total number of questions, examples include IELTS and TOFELS. Nowadays, standardization have evolved into different forms basing on these two prototypes. The standardization tests mentioned in this essay includes all formats used worldwide; the evaluation is not made toward a specific type.

3. Evaluation on the advantages of standardization tests

As mentioned, different types of standardization tests are used worldwide to assess student ability. Thus, various impacts are made by this large-scale implementation. Firstly, analysis on merits of standardization tests would be mentioned. The unique advantages occupied by this tool made it irreplaceable yet for many purposes.

3.1 Provide guidelines to be followed

The nature of standardization, being objective and standardized, turned the curriculum into an important guideline that supports both internally within the classroom and externally when adjusting curriculum.

3.1.1. In classroom to teachers

As criteria and tested contents are specially written in standardization tests, it is argued by Darling-Hammond [7] that students who takes these curricula receives a qualified education. It ensures teachers to cover a wide range of topics for the benefit of student, rather than distributing time unequally on areas teachers personally regard as more important. Thus, a more holistic education is created so by the time children are about to leave the school, they are equipped with comprehensive skill and knowledge.

Due to its relative objectivity, students with lower academic ability could be identified as well [6]. This qualitative data allows across comparison between students, revealing students that struggles with the content and needs support, which according to Shepard [8], is one of the purposes of testing. This numerical score also helps teachers to decide teaching schedules to meet specific student needs.

3.1.2. External support to other educators

Similarly, the external examiner could also follow the standard mark scheme provide by standardization tests to reduce examiner bias. All candidates are offered the same opportunity and assessed with the same criteria and standard. According to Weiss [9] and its replication of Birkel et al. [10], teachers are not always objective in terms of assessing student ability, especially without a fixed criterion. Thus, in-school test data like this that originally aimed to provide predications for student future performances might be as invalid and inaccurate. For example, college entrance tests. Standardization tests, on the other side, are usually designed to have standard answers and moderated between teachers. They are considered to provide more reliable and valid data that is more responsible for the student's future. [11]

3.2 Curriculum changes could be adapted quicker

Standardization tests also help curriculum contents being changed and adapted with a higher efficiency. Schools has always been modulated by governments and this influence is mainly reflected on the curriculum. According to Riffert [11], this is called an "in-put" orientated approach which manifests a top-down effect from the Ministry of Education to individual school and teachers. However, in the context of content changing, it is argued that this top-down influences are (a) slow and inflexible, (b) information of new contents are easily altered along the transformation process. The implemented strategies form individual schools and teachers sometimes could not meet the original spirit of the changed content. There exists a necessity for a tool that could adapt these changes

quicker, especially under this era where technologies and science innovation are tremendous. Standardized testing as an “out-put” approach, instead, makes the process explicit and visible, while the system only need to be changed according to the quality of the output. Thus, standardized testing is an ideal approach to achieve effective and fair school system.

4. Evaluations on disadvantages of standardization tests

Assorting comments and opinions around standardized testing, there are much more criticisms on the effect of its implementation comparing to the appraisal.

4.1 Not a valid indication of student ability

One of the most disapproved drawbacks of standardization tests is that it may not a valid indication of student ability. Test validity concerns whether what the test actually measures fit with the intention of what its developers want to measure, which is the student’s true academic ability. But the examinee’s on the spot performance could be influenced by many reasons: internal ones and external ones. For example, the stereotypical threat brought by the tests and the availability of teaching resources. Good et al. [12] argues that many standardized tests such as the SAT and the GRE evokes stereotypical threat by meeting two conditions: (a) Evaluative scrutiny in stereotypic domain and feeling pressured under unpleasant educational process, (b) group comparison especially related to gender and races when siting examinations with a mix of people . These conditions are often achieved by detailed identity checking before the tests and being placed in examinations rooms where the candidates naturally make group comparisons as there are a mix of people. As said by Inzlicht et al. [13] , groups composition could salient one’s social identity and stereotypes. The involvement of stereotypic threats could thus provide psychological implications that influence students’ spot performances [14]. As standardization tests have fixed process and results presented in numerical scores, many people consider it to have great objectivity, but Zitlow & Kohn [15] argues otherwise. Measuring student ability are not like changing independent variables in experiment, the manually written questions could be interpreted differently that leads to different understanding despite academic ability.

Moreover, curriculum validity as argued by Whitehead (1929/1967) [11], is not satisfied as well. Curriculum validity refers to the context that students have a fair chance to score on the test by practicing and learning the tested contents before hands. This is difficult to achieve especially on an international level due to high degree of diversification in education systems from different countries, which disadvantage students who are not taught according to the concept. To obtain curriculum validity, the teaching syllabus of test are usually nailed. But that takes away the freedom from teacher to adapt the courses toward student needs, being another obvious dilemma.

4.2 Teaching to the tests

Strictly stick with the tested content only in standardization tests leads to teaching for the tests. As suggested by Darling-Hammond [7], 60% of the teachers report that increased emphasis on standardization tests has affected their own teaching, and 95% report that this has affected other teachers' behaviours as well, namely in five ways: altered curriculum emphasis, test skill training, test preparation, having less time to reach to student and feeling under pressure.

4.2.1. Individual influences: Familiarization of test only

Many standardized tests are high-stake ones nowadays, they decide important academic milestone such as college entrance. As a result, pupils usually would just aim to pass or score highly on the test itself instead of understanding the tested concepts. The formula, bottom line and revision questions are much more valued. It turned to a skill of recognition and repetitive practice and memorizing. The interactive learning process is avoided, and it turned to a process of skill training. In the end, the score could increase. But that is simply due to students are more familiar with the question types. They do

not assess conceptual understanding, but instead value how much knowledge student could cram into their short-term memory.

Also, some exam boards profits are worsening this issue. While selling preparation materials designed to raise scores, children are often attracted to leave the school and study these specific materials. Education, according to Whitehead (1929/1967), is a complex process in which a number of factors play a crucial role, all of which must fit together in order to achieve the best result obtainable [11]. The best procedure of teaching will depend on several factors such as students' prospect of life, opportunities provided by the school, the involvement of parents but all these prospects could not be achieved as standardization only focus on the scores.

4.2.2. External influences: Critical thinking skills are not trained

A statical correlation between high scores on standardized tests and relatively shallow thinking have been found. Within the limited time given, the most important task to accomplish is to get as much correct answer possible rather than discovering the reasons behind. In comparison to spend time discover unique and creative answers, it is much safer for the majority of student to remember an answering format suitable for all. Even if it is a reading comprehension questions that "there are a thousand Hamlets in a thousand people's eyes", meaning each people could have understand it differently. If the exam strategy is smart enough, they could even just read the question then skim read the test to find the corresponding answer. As Lauren Resnick, one of the leading cognitive scientists put it: getting the right answer does not ensure understanding and getting the wrong one does not show the absence of understanding, but most standardized tests only care about the answer rather than the process of how you got it. Opportunists to practice important abilities such as critical thinking are not offered. "Focusing on this excruciatingly boring material would only fails to judge student capacity and drive away potential excellent scientists" said by President of National Academy of Sciences [15].

Famous educator Bill Ayers said "Standardized tests cannot measure imitative, creativity, imagination, conceptual thinking.... What they measure is the least significant part of education: the isolated skills and specific facts." Zitlow & Kohn [15] questions that these memorized facts and skill would not be helpful for the student future "How many jobs demand employees to come up with the correct, memorized answer within a few seconds without the help of others? How common is it for you to still deliver a pencil and paper exam when you already entered career life? I'd say only a few jobs need to, but the majority do not". Even students that does not plan to apply to a higher degree of education, they are still influenced by the implementation of standardization tests. They could have utilized this test revision and preparation time more wisely such as learning a practical skill that is more beneficial to him/her future.

4.3 Suppress of flexibility and creativity

Despite being criticized about its validity, standardization tests also do not count in individual uniqueness and interests. Whitehead (1929/1967) argues that people are unique and good teachers should adapt to these special capacities. But standardization, on the other hand, are unable to meet this, but assessing everyone from one perspective. [11]

4.3.1. Avoidance of personal talents leads to de-emphasis of subjects

Standardized tests limit the freedom of teachers to adapt to the complex, situation specific circumstances of individual students in order to obtain a maximum of the creative learning process for students who are gifted and interested in certain area of subjects. As a consequence, the attitude of creative, adventurous exploration is undermined and substituted by simple pattern recognition, narrow visions, and even boredom, neglecting the border scope of education.

Sarcastically, time spent on these boring practices usually substituted the time originally spent on areas and activities children are interested in. This forced learning mindset further creates negative toward studying, decreasing students' passivity, performance and vision, as suggested by Riffert (2005).[11] Carnoy et al. (2003) claims that these phenomena could potentially lead to de-emphasize

of certain untested subjects.[16] This not only suppress of creativity of the next generation. It is also worth questioning if children who went through standardized curriculum would all master tested subjects while lacking skills on those are not.

4.3.2. Creativity mindsets are not set

Pratzner, F. C. (1984), a Commerce Department study from the Office of Productivity, Technology, and Innovation, goes further in claiming that innovation and creativity are being squelched by "the basic educational philosophy," which claims standardization tests are better at "preserving convention knowledge than sparkling new invention." [7] Students are not practiced solving and develop logical and conceptual thinking to solve problems and take risk. Instead, it is argued that test-based knowledge is rival to practical knowledge that could solve problems in reality [17-18].

Whitehead (1929/1967) quotes it "The evocation of curiosity, of judgement, of the power of mastering a complicated tangle of circumstances, the use of theory giving foresight in special cases – all these powers are not to be imparted by a set rule embodied in one schedule of examination" [11]. Jerome Bruner [19] made the same argument half century later, "A method of instruction should have the objective of leading the child to discover for himself".

Moreover, it is worth mentioning that according to the data of National Assessment of Education Progress (NAEP), it was found that increase scores of basic reading and math was counterbalanced by decrease of more complicated subjects such as science and mathematical problem solving (National Research Council 1979). Some experts blame this decrease on the striving wish to achieve high score in standardized test. "What can be most easily tested and taught are now the teaching objectives in many schools," observed the National Council of Teachers of Mathematics president in hearings before the House Subcommittee on Elementary, Secondary, and Vocational Education. Teach children the easiest does achieves the goal but losing the further purpose of education-comprehension and application of knowledge [7].

4.4 Unnecessary stress and competitions are created

Unhealthy competitions are also promoted by the usage of standardization tests. This exists not only between students, but also in different schools and regions.

4.4.1. Individual influences: stress for students

As mentioned, Robert Glaser created the term 'criterion-referenced tests', meaning a type of assessment that tells one student's relative position in comparison to his/her peer who took the same assessment. These types of evaluation ignored student's own ability but instead grouped student into "successor who received better scores than the majority" and "failure who scored relatively lower" [15]. Being forced to compete with each other, student might not see their peers as friends anymore but competitors. In NTRs, there would always a small percentage that falls on the bottom in terms of "competitive percentage". Confidence of students could be diminished by these straight-forward and demotivating numbers, and slowly progressing a negative attitude towards learning.

Stress, as a result, are naturally created as a product of these competitions. "Emotions of suspicion, envy, self-doubt and hostility are evoked" commented by Zitlow & Kohn [15]. Test anxiety is already a subfield of educational psychology, meaning there exists a demand for children to learn how to cope with the pressure of standardization tests, especially when many of them are high-stake ones. Teachers, in the context of rising competition in the standardization tests, also experience pressure.

4.4.2. External influences: stress for educators

Meanwhile, unhealthy competitions between schools are also provoked. The tests cores are usually published regionally with the school's name to attract enrollment. In the era of big data, data such as standardization scores could be used in any places even though they were never intended to be used. If teaching efficiency is measure using the scores, all parents would like their children to be enrolled in the "best" school so as the teacher. Teaching resources and staffing are already imbalanced regionally. This further promoted the competition between states, province or even countries, causing

some students to move across places for schooling, which disadvantage students who are in states with less resources.

Being forced to implement test-centered curriculum decided by state legislators in priority to what is the most suitable for the children, many good teachers are driven out of this profession as a result. Not only the cocurricular activity are limited, but also field trips and school assemblies that offer children to learn important values outside of the book such as empathy. Kathy Greeley is a middle school teacher in Massachusetts. She designed a unique system that every student in her class have to select a unique area such as baking and ballet, then teach it to the rest of the class. This gives children a feeling of expertise in their own area, developing their confidence and communication ability. But now, Kathy Greeley claimed that "Because we have so much content material to cover, we do not have time to do it anymore". If standardization tests score are seen as a way to measure teaching effectiveness, then chances are low academic ability students are seen as obstacles and burden for teachers rather than opportunity for changes [15].

5. Suggestions

As standardization tests are argued to suppress children's creativity and talents. Many exam boards, such as the international baccalaureate categorized its subjects into six types, allowing students to choose for their own interests. According to Riffert [11], evaluations is necessary, but schools should adapt it suitably for their students. The Module Approach to Self-Evaluation of School Development Projects (MSS) is mentioned as an example to achieve the goal. This MSS module is developed by Dr. Andreas with his colleagues at the University of Salzburg, Australia. The national ministry of education in Australia offered the right for individual schools to change 1/3 of the contents of their curriculum according to students needs as suggested by the MSS. The other 2/3 of the teaching content would still be prescribed to the obligatory curriculum to ensure a standard education for students. Through this way, students could obtain the freedom to approach what interest and maximize their talents while relatively qualified education is ensured. However, the altered contents still need an objective and qualified measurement in order to evaluate students' ability as the curriculum is still within the range of standardization.

6. limitations

While lots of research are being done within the topic of standardization test's influences, the majority of available papers evaluate standardization tests holistically, therefore for research gaps are created. Firstly, most of research had taken the form of surveys and interviews. Although macro statistics and data are provided as prerequisites, these research methods rely on subjective opinions by participants and could exist bias. More objective and measurable experimental methods such as lab conditions should be encouraged to establish more specific cause and effect relationships. Secondly, the tendency of available papers to evaluate standardization tests holistically avoided differences between various tests and their question types. Evaluations are generally and unspecific, rather than pointing out advantages and disadvantages of each one. Thirdly, different subjects' performance in standardization testing is also waited to be discovered. As subjects' nature varies, could certain subjects to be considered more suitable for assessments than others? For instance, comparing question types and the layout of math examinations to reading comprehension tests. Lastly, student's age group and the relationship of how standardized could influence them is also worth considering. Children in different ages experiences various forms of academic expectations; generally speaking, this could change their attitude and impression towards the standardization tests.

7. Conclusion

While standardization tests are still being widely used, it remains controversial whether its indication and predication about students' academic ability are accurate, especially when they are

used to decide important milestones such as college admission. As this study aims to analyze and summarize both positive and negative of it. It could be concluded that on one hand, standardization tests provide quantitative data that allows comparison to be made between students; while on the other hand it could potentially leads to teaching to the tests and suppress individual uniqueness. Currently, there are no other suitable tools that could replace standardization tests with the same objectivity and qualification. Fortunately, however, the pros and cons of standardization tests are noticed by examination boards nowadays due to the research being done around this topic. Some curriculums are improved and incorporated parts where student could study according to individual needs.

References

- [1] M. Goertz, & M. Duffy. Assessment and Accountability Systems in the 50 States: 1999-2000 (2001). Retrieved 30 July 2022, from https://repository.upenn.edu/cpre_researchreports/13/
- [2] C. Jencks, & M. Phillips. The Black–White test score gap, Washington, DC: Brookings Institution Press (1998).
- [3] Lee J. Cronbach. Essentials of Psychological Testing, (4th ed.), New York: Harper and Raw Publishers Inc (1984).
- [4] T. Oakland, S. Douglas, and H. Kane,. Top Ten Standardized Tests Used Internationally With Children and Youth by School Psychologists in 64 Countries. *Journal of Psychoeducational Assessment*, 34(2), (2015) pp.166-176
- [5] T. Burkett. Norm-Referenced Testing and Criterion-Referenced Testing. *The TESOL Encyclopedia Of English Language Teaching*, 1-5 (2018). doi: 10.1002/9781118784235.eelt0351
- [6] R. Glaser. Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18(8), (1963) 519–521. <https://doi.org/10.1037/h0049294>
- [7] L. Darling-Hammond, & A. Wise. Beyond Standardization: State Standards and School Improvement. *The Elementary School Journal*, 85(3), (1985) 315-336. doi: 10.1086/461408
- [8] L. Shepard. The Role of Assessment in a Learning Culture. *Educational Researcher*, 29(7), (2000) 4-14. doi: 10.3102/0013189x029007004
- [9] R. Weiss. Die zuverlässigkeit der ziffernbenotung bei aufsätzen und rechenarbeiten. In R. Weiss, *Zensur und zeugnis*. Linz: Haslinger (1965).
- [10] P. Birkel, & C. Birkel. How Concordant Are Teachers' Essay Scorings?: A Replication Of Rudolf Weiss' Studies (2002). Retrieved from https://www.researchgate.net/publication/290771470_How_concordant_are_teachers%27_essay_scorings_A_replication_of_Rudolf_Weiss%27_studies
- [11] F. Riffert. The Use and Misuse of Standardized Testing: A Whiteheadian Point of View. *Interchange*, 36(1-2), (2005) 231-252. doi: 10.1007/s10780-005-2360-0
- [12] C. Good, J. Aronson, & M. Inzlicht. Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal Of Applied Developmental Psychology*, 24(6), (2003) 645-662. doi: 10.1016/j.appdev.2003.09.002
- [13] M. Inzlicht , & T. Ben-Zeev. A Threatening Intellectual Environment: Why Females Are Susceptible to Experiencing Problem-Solving Deficits in the Presence of Males. *Psychological Science*, 11(5), (2000) 365-371. doi: 10.1111/1467-9280.00272
- [14] C. Steele. A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), (1997) 613-629. doi: 10.1037/0003-066x.52.6.613
- [15] C. Zitlow, & A. Kohn. The Case against Standardized Testing: Raising the Scores, Ruining the Schools. *The English Journal*, 91(1), (2001) 112. doi: 10.2307/821673
- [16] Carnoy, Elmore, & Siskin. *New Accountability*. New York: Routledge (2003).
- [17] A. M. Bussis. “Burn It at the Casket”: Research, Reading Instruction, and Children’s Learning of the First R. *The Phi Delta Kappan*, 64(4), (1982) 237–241. <http://www.jstor.org/stable/20386656>

- [18] C. Kamii. Encouraging Thinking in Mathematics. The Phi Delta Kappan, 64(4), (1982) 247–251.
<http://www.jstor.org/stable/20386658>.
- [19] J. Bruner. The relevance of education. W. W. Norton & Company (1971).